BU-283-M

SELECTION MODELS FOR A QUANTITATIVE TRAIT IN NEUROSPORA*

W. T. Federer, K. E. Papa**, and A. M. Srb

ABSTRACT

Five statistical forms of a model for increase in the generation mean value of a population under selection were presented and discussed. Procedures were advanced on methods of estimating the parameters of the statistical models and of fitting the data to a theoretical response curve. A limited study on empirical and theoretical properties of the estimators was conducted. It was possible to find least squares estimators which were conditional upon the ratio of the estimated environmental to genotypic variances being the true ratio. Variances for the conditional least squares estimators follow directly from regression theory.

A genetic example of an intra-strain cross of Neurospora, Honla/Hon3A, and its reciprocal was used to illustrate the procedures. Each cross was replicated twice at each of three temperature levels (18°C., 25°C., and 35°C.), and the character studied was linear growth rate in growth tubes. One of the statistical models, not involving the normality assumption, resulted in a good fit for all 12 sets of data up to 11 and 12 cycles, or generations, of selection. (Other data up to 20 generations also resulted in good fits to the postulated statistical model.)

The logarithmic and square root transformations of the data tended to stabilize, but not equalize, the variances for the various temperature levels. The transformations had no apparent effect on the properties of the estimators in that an empirical investigation of values of the parameters resulting in a minimum sum of squares, was relatively unaffected by the transformation.

---

* Paper No. BU-96 of the Biometrics Unit, and paper No. 464 of the Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.

** Presently located at the University of Georgia, Athens, Georgia.

SELECTION MODELS FOR A QUANTITATIVE TRAIT IN NEUROSPORA*

W. T. Federer, K. E. Papa**, and A. M. Srb

1. INTRODUCTION

Since advance in the mean value of a character under selection depends upon the residual genetic variance from the preceding generation and upon the selection pressure, a simple recursive model may be used to describe the population mean at the i^{th} step in a selection program. Five forms of such a model are presented together with estimators and variances for one of them. In addition, an application is made to one set of data obtained from the experiments of Papa [1964]. An empirical examination of the effect of transformations and of properties of estimators is made.

The genetic basis for one or another of the forms discussed herein has been discussed in a number of places in published literature. (e.g. Falconer [1960]; Kempthorne [1952] and references in Chapter 23; etc.) Results from long-term selection experiments have appeared in genetic literature from time to time (e.g. see Falconer [1960], Ch. 12). However, the problems of fitting a curve (other than straight line) to a set of selection data and of estimating the associated parameters appear to have received no treatment in the literature. Hence, the purpose here is to introduce a discussion of these problems.

For the models described herein, it is assumed that a large number of factors affect the character under consideration and that the factors affecting

---

* Paper No. BU-96 of the Biometrics Unit, and paper No. 464 of the Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.

** Presently located at the University of Georgia, Athens, Georgia.

a character are similar in expression. (If few factors control the expression of the character under consideration, then some such model as described by Federer, Robson, and Srb [1959] would suffice.) The genotypic effects are assumed to be random variables identically and independently distributed with mean zero and common variance σ_g^2. The environmental effects are also assumed to be identically and independently distributed random variables with zero mean and common variance σ_e^2. In the bivariate distribution of environmental and genotypic effects, a zero covariance is postulated. This implies no gentoype x environment interactions for the environments encountered in the collection of a set of data. Whether or not this assumption is justifiable depends upon the genetic material and the environments encountered.

2. SELECTION MODELS

We shall consider the particular case of selection in which the phenotypic effect is the sum of the genotypic effect and the environmental effect, and the phenotypic variance, σ_p^2, is the sum of the genotypic variance, $\sigma_g^2 \geq 0$, and environmental variance, $\sigma_e^2 > 0$, i.e. $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$ (see Falconer [1960], chapter 11, and Kempthorne [1957], chapter 23). Further let us define the difference between the mean of the selected portion of the population and the mean of the unselected population to be S_i, the selection differential in the i^{th} generation. Since individuals are selected on the basis of their phenotype and since the regression of genotype on phenotype is $b_{i-1}\sigma_g^2/(\sigma_e^2 + b_{i-1}\sigma_g^2)$, where b_{i-1} reflects the residual heterozygosity due to the breeding system (e.g., $b_{i-1} = 1/2^{i-1}$ in the selfing series), the expected response due to selection in the $i+1^{st}$ generation is

$$E(R_{i+1}) = E(\bar{y}_{i+1,\cdot} - \bar{y}_{i\cdot}) = \frac{b_i \sigma_g^2 S_{i+1}}{b_i \sigma_g^2 + \sigma_e^2} \tag{2.1}$$

$$= \frac{b_i S_{i+1}}{b_i + \beta} = \frac{b_i H S_{i+1}}{1 + H(b_i - 1)}$$

where $\beta = \sigma_e^2/\sigma_g^2$, $H = \sigma_g^2/(\sigma_e^2+\sigma_g^2)$ = heritability (see Lush [1945]), and $\bar{y}_{i+1,\cdot}$ and $\bar{y}_{i\cdot}$ are the sample or line means in the $i+1^{st}$ and i^{th} generations, respectively. It should be noted that $E(R_{i+1})$ depends only upon the ratio $\beta = \sigma_e^2/\sigma_g^2$ and not on the two parameters σ_g^2 and σ_e^2 individually. When the distribution of the phenotypic effects is in the family of distributions in which the scale parameter is proportional to the standard deviation we may write equation (2.1) as:

$$E(R_{i+1}) = \frac{b_i \sigma_g^2 S_{i+1}}{b_i \sigma_g^2 + \sigma_e^2} = \frac{b_i \sigma_g^2 \left(\sqrt{b_i \sigma_g^2 + \sigma_e^2}\ \bar{x}_{(mi)} a_{i+1}\right)}{b_i \sigma_g^2 + \sigma_e^2} \tag{2.2}$$

$$= \frac{b_i \sigma_g \left(\sigma_g \bar{x}_{(mi)} a_{i+1}\right)}{\sqrt{b_i \sigma_g^2 + \sigma_e^2}} = \frac{b_i \delta_i a_{i+1}}{\sqrt{b_i + \beta}}$$

where a_{i+1} is the selection pressure in the $i+1^{st}$ generation and where $\delta_i = \sigma_g \bar{x}_{(mi)}$ = the genetic standard deviation times the expected value of the largest (or the mean of the k largest) order statistic, $\bar{x}_{(mi)}$, for the particular standardized distribution under consideration for a sample of m selections in generation i. (See Sarhan and Greenberg [1962], e.g.) In this form, we

need only know that the form of the distribution from this general family of distributions remains unchanged, that the number of selections (mi) remains unchanged (i.e., mi = m), that the selection pressure remains unchanged (i.e. $a_{i+1} = 1$) and that the genetic variance σ_g^2 remains unchanged, then the value of δa_{i+1} is a constant. If selection pressure changes we let a_{i+1} in equation (2.2) reflect this and also account for change in m. The normal distribution is a special case of this family of distributions for which the scale parameter is the standard deviation. Values of $\bar{x}_{(m)}$ for the largest, second largest, etc. order statistics from a sample of size m from a normal distribution have been tabulated in a number of places (e.g., see Fisher and Yates [1938], Table XX; Federer [1951], Tables 3 and 4; and Harter [1961], Table 1).

Suppose now that the unselected population mean is α and that selection is practiced. Then, from equations (2.1) and (2.2) the linear equation representing the observed value of the j^{th} observation in the i^{th} generation is

$$Y_{ij} = \alpha + \sigma_g^2 \sum_{h=0}^{i} \frac{b_{h-1} S_h}{b_{h-1}\sigma_g^2 + \sigma_e^2} + \epsilon_{ij}$$

$$= \alpha + H \sum_{h=0}^{i} \frac{b_{h-1} S_h}{1 + (b_{h-1})H} + \epsilon_{ij}$$

$$= \alpha + \sum_{h=0}^{i} \frac{b_{h-1} S_h}{b_{h-1} + \beta} + \epsilon_{ij}$$

$$= \alpha + \sigma_g^2 \bar{x}_m \sum_{h=0}^{i} \frac{b_{h-1} a_h}{\sqrt{b_{h-1}\sigma_g^2 + \sigma_e^2}} + \epsilon_{ij}$$

$$= \alpha + \delta \sum_{h=0}^{i} \frac{b_{h-1} a_h}{\sqrt{b_{h-1} + \beta}} + \epsilon_{ij} \qquad (2.3)$$

where the ϵ_{ij} are random variates identically and independently distributed with mean zero and common error variance σ_e^2, $b_{-1} = 0$, $b_{i-1} = 2^{i-1}$ (or some other value), $i=0,1,2,\cdots,n$ generations, $j=1,2,\cdots,r_i$ replicates per generation, and the other symbols are as defined above. The particular form of the genetic model appropriate in equations (2.3) depends upon the knowledge available in the experiment. In certain situations the selection differential S_i and the response to selection R_i are known. In this case the estimated heritability in generation i is simply R_i/S_i (see Falconer [1960], Chapter 11). Also, if values of R_i and estimated values for S_i are available from n generations an estimate of the parameter β could be obtained from the solution of the following equation for $\hat{\beta}$:

$$\sum_{i=0}^{n-1} R_{i+1} = \sum_{i=0}^{n-1} (\bar{y}_{i+1,\cdot} - \bar{y}_{i\cdot}) = \sum_{i=0}^{n-1} b_i S_{i+1} / (b_i + \hat{\beta}) \quad . \qquad (2.4)$$

For the particular experiment to be considered herein, the last form of the five forms in equations (2.3) was deemed appropriate. Also, the $\Sigma\Sigma\epsilon_{ij}^2$ will not always reach a minimum for simultaneous estimation of α, δ, and β (see section 5). Whether or not a solution for the three parameters is possible

depends upon the values of the Y_{ij} and the coefficients b_i. Since this is true a separate independent estimate of β was obtained from associated experiments in this study. Then, the usual least squares estimates of α and δ hold, conditional upon the fact that $\hat{\beta} = \beta$. The conditional least squares estimates of α and δ are obtained by differentiating $\Sigma\Sigma\epsilon_{ij}^2$ with respect to α and to δ, by setting the resulting equations equal to zero, and by solving the two resulting equations for α and δ. The conditional least squares estimates of α and δ are:

$$\hat{\alpha} = \bar{y} - \hat{\delta} \sum_{i=0}^{n} r_i W_i / r_. \tag{2.5}$$

and

$$\hat{\delta} = \frac{\sum_{i=0}^{n} W_i (Y_{i.} - r_i \bar{y})}{\sum_{i=0}^{n} r_i W_i^2 - (\Sigma r_i W_i)^2 / r_.} \tag{2.6}$$

where $W_i = \sum_{h=0}^{i} b_{h-1} a_h / \sqrt{b_{h-1} + \beta}$, $Y_{i.} = \sum_{j=1}^{r_i} Y_{ij}$, $\bar{y} = \sum_{i=0}^{c} \sum_{j=1}^{r_i} Y_{ij} / r_.$, and $r_. = \sum_{i=0}^{c} r_i$. The variances of $\hat{\alpha}$ and $\hat{\delta}$ are the usual variances for the intercept and slope, respectively, for linear regression.

As stated above estimates of σ_e^2 and σ_g^2, or their ratio $\sigma_e^2/\sigma_g^2 = \beta$, is required for the above estimates $\hat{\alpha}$ and $\hat{\delta}$. The meaning of the statement that the ϵ_{ij} have common variance σ_e^2 requires amplification. Basically, this statement implies that for a true population with no genetic changes, the expected value of the variance among n observations taken singly in each of the environments in which the $\bar{y}_{i.}$ are obtained is equal to the expected value of the variance

among s observations in generation i. In the analysis of variance table for s observations in each generation, this would mean:

Source of variance	d.f.	Average value of mean square
Among generations	n	σ_e^2
Within generations	(n+1)(s-1)	σ_e^2

Also, this implication could be expressed symbolically as:

$$V(Y_{ij}|i) = V(Y_{ij}) = E(Y_{ij}-EY_{ij})^2 = \sigma_e^2 \quad .$$

If the progeny from the different generations of selections are all compared in one experiment, estimates of α, σ_e^2, and σ_g^2 may be obtained from s individuals (spores, strains, etc.) from the unselected population each replicated r times. Then $\hat{\alpha} = \sum_{h=1}^{s} \sum_{j=1}^{r} Y_{ij}/rs = \bar{y}$ and $\hat{\sigma}_e^2$ and $\hat{\sigma}_g^2$ may be obtained from the following analysis of variance on the unselected individuals:

Source of variation	d.f.	Mean square	E[m.s.]
Among individuals	s-1	A	$\sigma_e^2 + r\sigma_g^2$
Within individuals	s(r-1)	B	σ_e^2

as $\hat{\sigma}_e^2 = B$ and $\hat{\sigma}_g^2 = (A-B)/r$.*

---

* There are several estimators for σ_e^2 and σ_g^2 in the literature, but a discussion of these is not pertinent to this paper (see e.g. Federer [1962]).

If the generation means are obtained from a series of experiments, the unselected population will necessarily be included in each experiment if the variance of observations among experiments is different from the variation among individuals within experiments. If the variance of observations among experiments is equal to the variance within experiments, then a single experiment could be conducted at the beginning of the selection program and estimates of α, σ_e^2, and σ_g^2 obtained at this stage.

For any of the models postulated by equations (2.3), or any variation of them, an estimator for σ_e^2 is simply the sum of squares of deviations between the generation mean and the fitted point on the curve divided by the degrees of freedom. Whether or not the variance among the r_i observations in generation i is an estimator for σ_e^2 depends upon the conditions stated in the preceding paragraph. For the data considered in the next section this is not the case. The variation among duplicates grown in one environment is considerably smaller than that from experiment to experiment. In this case the variation among generation means of a single unselected population with no genetic change is $\sigma_e^2 + \sigma_d^2/d$ where σ_d^2 is the variance among duplicates in one environment and d is the number of d-tuplicates grown in the one environment. Also, for the data given here σ_g^2 estimated from the data from n generations poses some problems. These are considered in section 4.

For β known, the ordinary least squares estimators for the intercept and the slope are estimators for α and δ. Since the $\sum_{h=0}^{\infty} \frac{b_{h-1}}{\sqrt{b_{h-1}+\beta}}$ = a constant = C_1 for $\beta > 0$, since the b_{h-1} are of the order of 2^{-h+1} for inbreeding, and since the h^{th} term of this series approaches zero as h increases, the estimators for

α and δ are not even consistent. That this is so can be observed from the variance of $\hat{\delta}$ given β where the denominator is of the form $\sum_{i=0}^{n} W_i^2 - \left(\sum_{i=0}^{n} W_i\right)^2/(n+1)$ for $W_i = \sum_{h=0}^{i} \frac{b_{h-1}}{\sqrt{b_{h-1}+\beta}}$. The W_i are ordered and rapidly approach the constant C_1. This means that the $\sum_{i=0}^{n} W_i^2 - \left(\sum_{i=0}^{n} W_i\right)^2/(n+1)$ does not become larger as n increases, but is always less than $\sum_{i=0}^{n} (W_i - C_1)^2$, which does not increase in value for a specified number of significant figures after i = some number N.

Therefore, in order to have consistent estimators for α and δ given β, it would be necessary to replicate experiments for a fixed number of generations, i.e., increase the r_i at the expense of the number of generations n. In fact, the first few observations, say generations 0, 1, and 2, are much more important generations for estimating the parameters α and δ than the later generations; after i=N a specified number, additional generations are essentially of no value in estimating α and δ. The more efficient statistical procedure must, of course, be viewed in light of biological considerations. One of the more important biological considerations is to determine if the postulated model fits for an "adequate" number of generations ("adequate" is defined here to mean until the biologist becomes tired of conducting experiments). Thus, from a statistical point of view the most efficient sampling procedure would be to use replicated observations from two generations, 0 and 1, to fit the model postulated by equation (2.3) for a specified β. From certain biological points of view it would appear that 10 to 15 or more generations would suffice to observe the appropriateness of the models postulated herein; certain types of experiments may require additional generations, say 30 to 100 generations; as a compromise between the statistical and biological

considerations it is suggested that r_i be relatively large for i=0 and 1 and be equal to unity for i=2,3,$\cdots$.

For models of the nature postulated by equations (2.3), careful thought must be given to the nature of the parameters being estimated in relation to the sampling plan and the observations. There appears to be a redundancy for some of the estimators obtained. For example, consider the following sum of squares:

$$\sum_{i=0}^{n} \sum_{j=1}^{r_i} \left[Y_{ij} - \alpha - \sigma_g^2 \sum_{h=0}^{i} \frac{b_{h-1} \bar{x}_{mh}}{\sqrt{b_{h-1}\sigma_g^2 + \sigma_e^2}} \right]^2$$

In the above there is a temptation to estimate α, σ_g^2, and σ_e^2. But, the above sum of squares divided by n and with the parameters replaced by estimates of parameters is used for estimating σ_e^2. Since σ_e^2 is contained inside the summation, it appears that an estimate of σ_e^2 must be obtained in another manner; then, σ_e^2 is replaced by its estimate and estimators for σ_g^2 and α are obtained.

3. A GENETIC EXAMPLE

A selection program originating from several inter- and intra-strain crosses of Neurospora, as well as their reciprocals, was carried out at three different temperatures, 18°, 25° and 35°C. (Papa [1964]). Selection for fast growth rate was practiced among vegetative cultures arising from single ascospores which were isolated at random from each cross. Ten cultures of each mating type, A and a, from any one cross were grown in duplicate growth tubes on minimal media at each temperature in each generation. Methods used for

preparing and handling growth tubes were those described by Ryan, et al., [1943]. The fastest growing cultures of opposite mating type were crossed to obtain the population for the next generation. Second samples taken from the original crosses and subsequent cycles of selection constituted a second replication. Standard techniques for crossing strains and making isolations were used (Beadle and Tatum [1945]).

Whenever less than ten individuals were obtained the a_h in equations (2.3) were computed from Table XX of Fisher and Yates [1938] as follows. The expected value of the largest member from a sample of size ten is 1.54, from a sample of size nine is 1.49, from a sample of size eight is 1.42, etc. If ten A and ten a individuals were available in generation h then $a_h = 1$. If nine were available for A and ten for a, then $a_h = (1.49 + 1.54)/2(1.54) = 2(1.49)/2(1.54) = .968$; if eight were available for A and nine for a then $a_h = (1.42 + 1.59)/2(1.54) = .945$; etc. This normal approximation for changes in selection pressure was utilized even though normality is not assumed in obtaining $\hat{\alpha}$ and $\hat{\delta}$ in equations (2.5) and (2.6).

To illustrate the procedure which was used for comparing experimental data with a theoretical model, a detailed example will be given for part of the data. The selection results obtained from crosses between opposite mating types of a wild-type strain of Neurospora collected in Honduras will be used for this purpose. Selection was initiated in the cross, Honla/Hon3A, and its reciprocal, Hon3A/Honla. The parent to the left of the diagonal line acted as the conidial parent and the one to the right as the protoperithecial parent.

Mean growth rates of the progeny from each cycle of selection for both replications of the two crosses at these temperatures are presented in Figure 1.

It should be noted that the reciprocal cross Honla/Hon3A appeared to yield higher growth rates by the time selection had been practiced for 10 to 11 generations. Whether or not this is sampling variation would need to be tested in further experimentation.

The analyses of variance for each cycle, replication, and temperature for both crosses are presented in Table 1. Occasionally some observations were missing but for the most part either 19 or 20 individuals were included in each cycle of these two crosses. Duplicate growth results were available for most of the individuals in each generation. In some cases for generation 11 for the cross Honla/Hon3A and for generation 10 for the cross Hon3A/Honla only single tubes were used to measure growth rate. Thus, there would be no within individuals mean square in these situations; no use was made of the among individuals mean square in computing an estimate of σ_g^2.

For the experimental conditions encountered the variation between duplicate growth tubes obtained at one time appears to be considerably different from duplicate growth tubes grown at different times. Therefore, the within individuals mean square is defined to have the expectation σ_d^2, a component of variance due to duplicate determinations obtained at one time and which is a different component of variance from that due to variation from time to time of experimentation. Since the degrees of freedom are essentially equal, the within mean squares are averaged to obtain the following average estimates of σ_d^2:

Honla/Hon3a			Hon3A/Honla		
18°	25°	35°	18°	25°	35°
.00152	.00401	.00746	.00291	.00677	.00441

The expectation of the among individuals mean squares is defined to be $\sigma_d^2 + 2^{-i}d\sigma_g^2$, where d is the number of growth tubes for each individual and 2^{-i} is the coefficient from an inbreeding series for generation i. From the analyses of variance for one temperature an estimator for σ_g^2 is obtained by minimizing the following sum of squares with respect to σ_g^2:

$$\sum_{j=1}^{r} \sum_{i=0}^{n} \left[(A_{fi}-U_{fi})/d - 2^{-i}\sigma_g^2 \right]^2$$

with the result

$$\hat{\sigma}_g^2 = \sum_{f=1}^{r} \sum_{i=0}^{n} (A_{fi}-U_{fi})2^{-i}/d \sum_{f=1}^{r} \sum_{i=0}^{n} 2^{-2i} \quad .$$

For the genetic example the estimates of $\hat{\sigma}_g^2$ for each replication (i.e., for only one replicate n=1 in the above formula) at 18°, 25°, and 35°C. are:

Temp.	Hon1a/Hon3A			Hon3A/Hon1a			Both crosses over two replicates
	Rep I	Rep II	I & II	Rep I	Rep II	I & II	
18°C.	.016	.079	.048	018	.011	.015	.031
25°C.	073	.117	.095	.023	.058	041	.068
35°C.	.255	.236	.245	.155	.143	.149	.197

The estimated genetic variance, $\hat{\sigma}_g^2$, for the cross Hon3A/Hon1a was lower than the corresponding averages of both crosses at each temperatures for both replicates. This is an apparent confirmation of the selection results in Figure 1 where the highest mean growth was obtained from the cross Hon1a/Hon3A.

From regression theory the estimated variance could be computed as follows:

$$\widehat{V(\hat{\sigma}^2_g)} = \sum_{f=1}^{r} \sum_{i=0}^{n} \left[A_{fi} - U_{fi} - 2^{-i} d \hat{\sigma}^2_g \right]^2 / rd^2(rn+r-1) \sum_{i=0}^{n} 2^{-2i} \quad .$$

Alternatively, σ^2_{gi} could be computed for the i^{th} generation from the data from that generation. The variable results for each generation indicate that considerably more selections, say 50 to 100, would be required in order to obtain stable estimates of σ^2_g in the i^{th} generation. The sampling variation in σ^2_g computed from results in each generation and utilized in equations (2.3) to (2.6) would produce a broken line curve for the fitted curve.

In all of the above it has been assumed that there is no change in the parameter σ^2_g other than that due to the breeding system, e.g inbreeding. Loss of genetic variation because of a finite population size could reduce σ^2_g further than given by $b_i \sigma^2_g$. Also, mutation and contaminants could increase the actual genetic variance. These sources could produce a bias and though hard to measure it should be considered whenever interpretations of a set of data such as these are being interpreted.

An among and within generation analysis of variance on the means of Figure 1 would lead to an analysis of variance yielding an estimate of the environmental variance described in section 2. The within generation mean square is an estimate of $\sigma^2_e + \sigma^2_d/sd$, where s = number of spores per generation (20 here) and d = number of duplicate growth tubes (2 here). That is, each replicate and generation mean arises from sd = 40 observations. Thus, an estimate of σ^2_e is obtained as the within generation mean square minus the estimated $\sigma^2_d/40$.

The within generation mean squares obtained for the genetic example were:

	Temperature		
	18°C.	25°C.	35°C.
Honla/Hon3A	.00487	.01343	.05142
Hon3A/Honla	.00629	.00651	.00952
Combined	00596	.02047	.03930

Since normality of environmental and genetic effects may be unrealistic and since the selection data were thought to follow the last form of equations (2.3), the parameters α, β, and δ were estimated and theoretical curves were fitted to the data for each of the three temperatures. Therefore, using $\hat{\sigma}^2_g$ as the estimate of σ^2_g, $\hat{\beta}$ may be computed as $(\widehat{\sigma^2_e+\sigma^2_d}/40)/\hat{\sigma}^2_g = .00487/.048 = .102$, which is the value for $\hat{\beta}$ the 18°C. used in the computations in Table 2.

Utilizing the formulae (2.5) and (2.6) and the independent estimate $\hat{\beta}$ from Table 1, the conditional least squares estimates may be obtained for the data of Figure 1. The results are described in Table 2 for the cross Honla/Hon3A at 18°C. The computed curves and the actual data are plotted in Figure 1 for the values of $\hat{\alpha}$, $\hat{\delta}$, and $\hat{\beta}$ listed thereon. The predicted growth rates for the 18°C., 25°C. and 35°C. data for each cross and for the crosses combined are given in Table 3 and are plotted in Figure 1.

4. TRANSFORMATION OF THE DATA

In order to obtain more precise estimates of σ_d^2 and σ_e^2 the various temperature level data could be combined if the variances were not heterogeneous. From the results given in the last section it is apparent that the variances increased, in general, with temperature level and with increased growth rate. Since the logarithmic and square root transformations are variance stabilizing transformations for data of the above described type, the data for all 14 crosses at three different temperatures were transformed by the square root and logarithmic transformations (see Papa [1964]). Although these transformations tended to stabilize variances more than for the untransformed data, they were not completely successful for all crosses. Hence, the less efficient procedure of handling each cross at each temperature level separately was used as described in the previous section.

Some crosses may exhibit different types of or more genotype $\times$ environmental interaction than others; hence, the above transformations may not stabilize the variances. More experimental information is required in order to be able to determine which function of the data results in stable variances over different temperature levels.

5. MINIMUM SUM OF SQUARES

An empirical investigation was conducted to determine if values of α, δ, and β existed which would make $\sum_{i=0}^{n} \sum_{j=1}^{r_i} \epsilon_{ij}^2$ from the last form of equations (2.3) a minimum. As a starting point the $\hat{\beta}$ values for the genetic example were utilized. Then, the following multiples of $\hat{\beta}$ were used in equations (2.5) and (2.6) to estimate α and δ:

0, 1/16, 1/8, 1/4, 1/2, 1, 3/2, 2, 3, 4, 6, 8, 12, 16. The sum of squares of the deviations of observed from predicted values was computed for each set of estimates, making a total of 81 (1 replicate by 3 temperatures by 1 cross and 2 replicates by 3 temperatures by 13 crosses) sets of sums of squares. The resulting sums of squares were plotted against multiples of $\hat{\beta}$ to determine the shape of the response curve and the place where a minimum occurred. The following summarizes the information:

Multiples of $\hat{\beta}$	Minimum sum of squares (frequency)
0	37
1/16	4
1/8	1
1/4	4
1/2	4
1	4
3/2	0
2	0
3	0
4	0
6	0
8	0
12	0
16	20

The remaining seven sets were of the following nature:

3 had a minimum at 16 and a maximum at 1/16

1 " " " " 0 " " " " 1/4

1 " " " " 0 " " " " 1/16

1 " " " " 0 " " " " 1

1 " " " " 16 " " " " 1/4

Thus, only 17 of the 81 sets had a minimum sum of squares in the positive part of the admissible range of β (i.e., non-negative values). About one out

of five sets of data were responding in the manner expected for least squares estimates. Also, it would appear that 20 of the 81 sets would attain a minimum sum of squares for $\beta = \infty$, since the minimum sum of squares was attained for $\beta = 16\hat{\beta}$ and since the curves appeared to be approaching an asymptotic limit for increasing β.

The 12 sets of mean squares for 14 different values of $\hat{\beta}$ for the genetic example considered herein, are presented in Table 4. The remaining 69 sets are given by Papa [1964]. The data in Table 4 exemplifies three of the types of responses described above with one of these, Rep II at 35°C. for Honla/Hon3A, responding as expected for least squares estimates.

For the experiments described by Papa [1964], the estimators for α and δ from equations (2.5) and (2.6) are least squares estimators conditional upon the fact that the independent estimate $\hat{\beta} = \beta$. In one out of five experiments it was possible to find values of α, δ, and β which minimized the residual sum of squares. Square root and logarithmic transformations of the data did not alter the above results to any degree.

6. SUMMARY

Five statistical forms of a model for increase in the generation mean value of a population under selection were presented and discussed. Procedures were advanced on methods of estimating the parameters of the statistical models and of fitting the data to a theoretical response curve. A limited study on empirical and theoretical properties of the estimators was conducted. It was possible to find least squares estimators which were conditional upon the ratio

of the estimated environmental to genotypic variances being the true ratio. Variances for the conditional least squares estimators follow directly from regression theory.

A genetic example of an intra-strain cross of Neurospora, Honla/Hon3A, and its reciprocal was used to illustrate the procedures. Each cross was replicated twice at each of three temperature levels (18°C., 25°C, and 35°C.), and the character studied was linear growth rate in growth tubes. One of the statistical models, not involving the normality assumption, resulted in a good fit for all 12 sets of data up to 11 and 12 cycles, or generations, of selection. (Other data up to 20 generations also resulted in good fits to the postulated statistical model.)

The logarithmic and square root transformations of the data tended to stabilize, but not equalize, the variances for the various temperature levels. The transformations had no apparent effect on the properties of the estimators in that an empirical investigation of values of the parameters resulting in a minimum sum of squares, was relatively unaffected by the transformation.

8. LITERATURE CITED

Beadle, G. W., and E. L. Tatum, 1945. Neurospora II. Methods of producing and detecting mutations concerned with nutritional requirements. American Journal of Botany 32:678-686.

Falconer, D. S., 1960. Introduction to quantitative genetics, Chapters 11, 12, and 13. Ronald Press Co., N. Y.

Federer, W. T., 1951. Evaluation of variance components from a group of experiments with multiple classifications. Iowa Agricultural Experiment Station Bulletin 380.

Federer, W. T., 1962. Estimators for components of variance. Biometrics Unit, Cornell University, Mimeograph Series BU-144-M.

Federer, W. T., D. S. Robson, and A. M. Srb, 1959. A model testing problem in genetics. Biometrics Unit, Cornell University, Mimeograph Series BU-103-M.

Fisher, R. A., and F. Yates, 1938. Statistical tables for biological, agricultural and medical research, 1st ed. Oliver and Boyd, Edinburgh.

Harter, H. L., 1961. Expected values of normal order statistics. Biometrika 48:151-165.

Kempthorne, O., 1957. An introduction to genetic statistics, Chapter 23. John Wiley and Sons, Inc., N.Y. and London.

Lush, J. L., 1945. Animal breeding plans. Iowa State College Press, Ames, Iowa.

Papa, K., 1964. Effectiveness of selection based on variability uncomplicated by heterotic effects in Neurospora. Ph.D. Thesis, Cornell University, September.

Ryan, F. J., G. W. Beadle, and E. L. Tatum, 1943. The tube method of measuring the growth rate of Neurospora. American Journal of Botany 30:784-799.

Sarhan, A. E., and B. G. Greenberg, 1962. Contributions to order statistics. John Wiley and Sons, Inc., N. Y. and London.

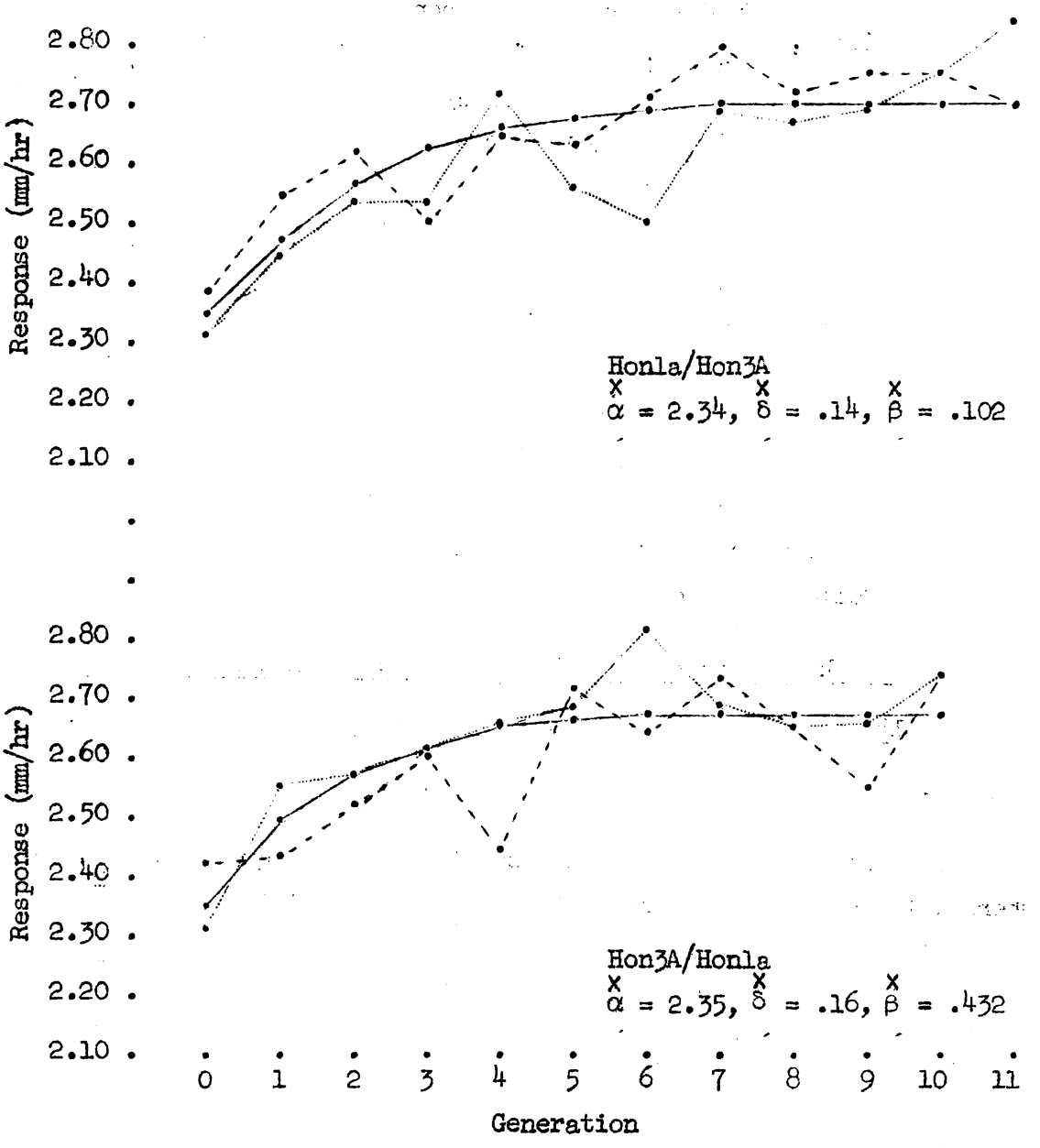
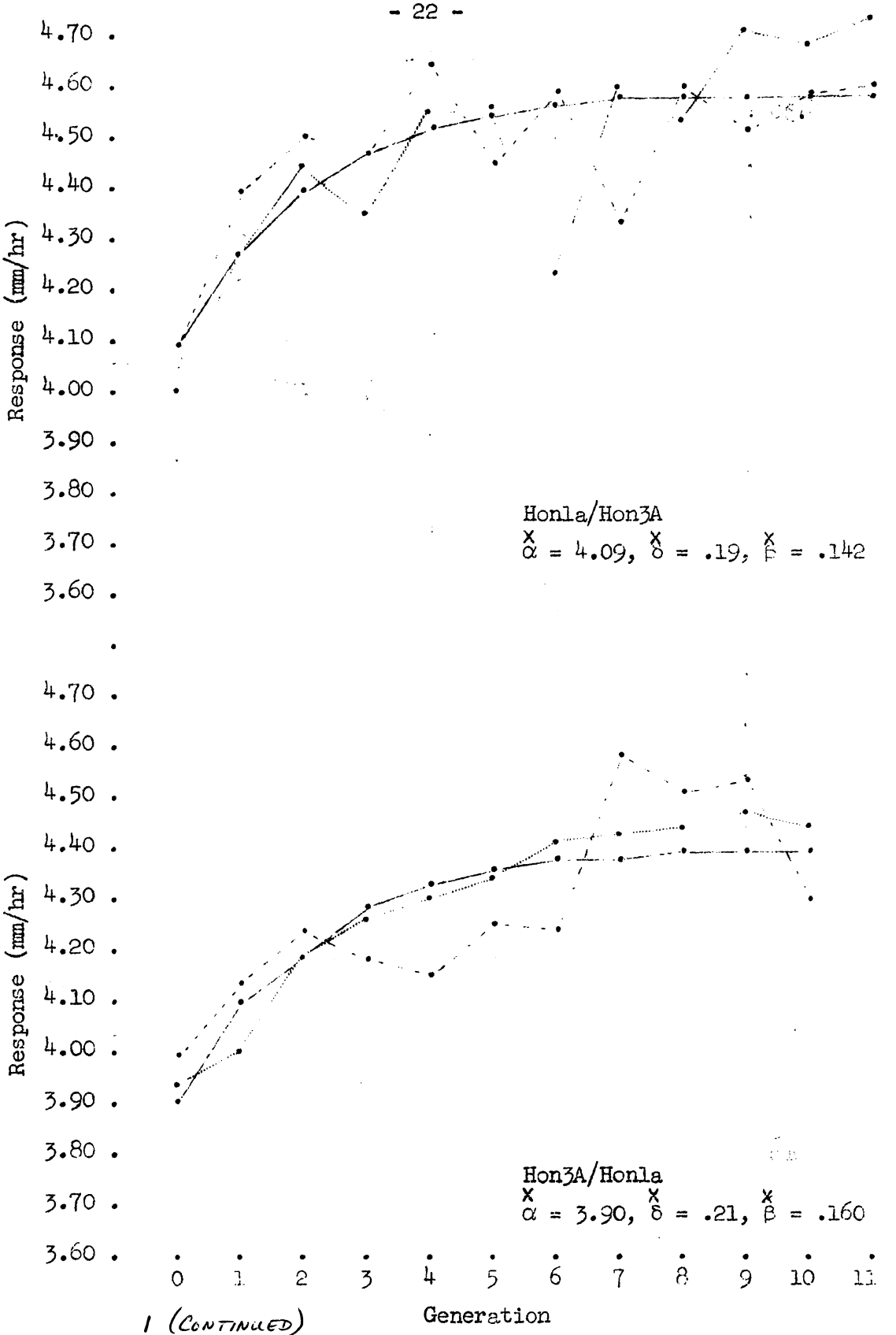
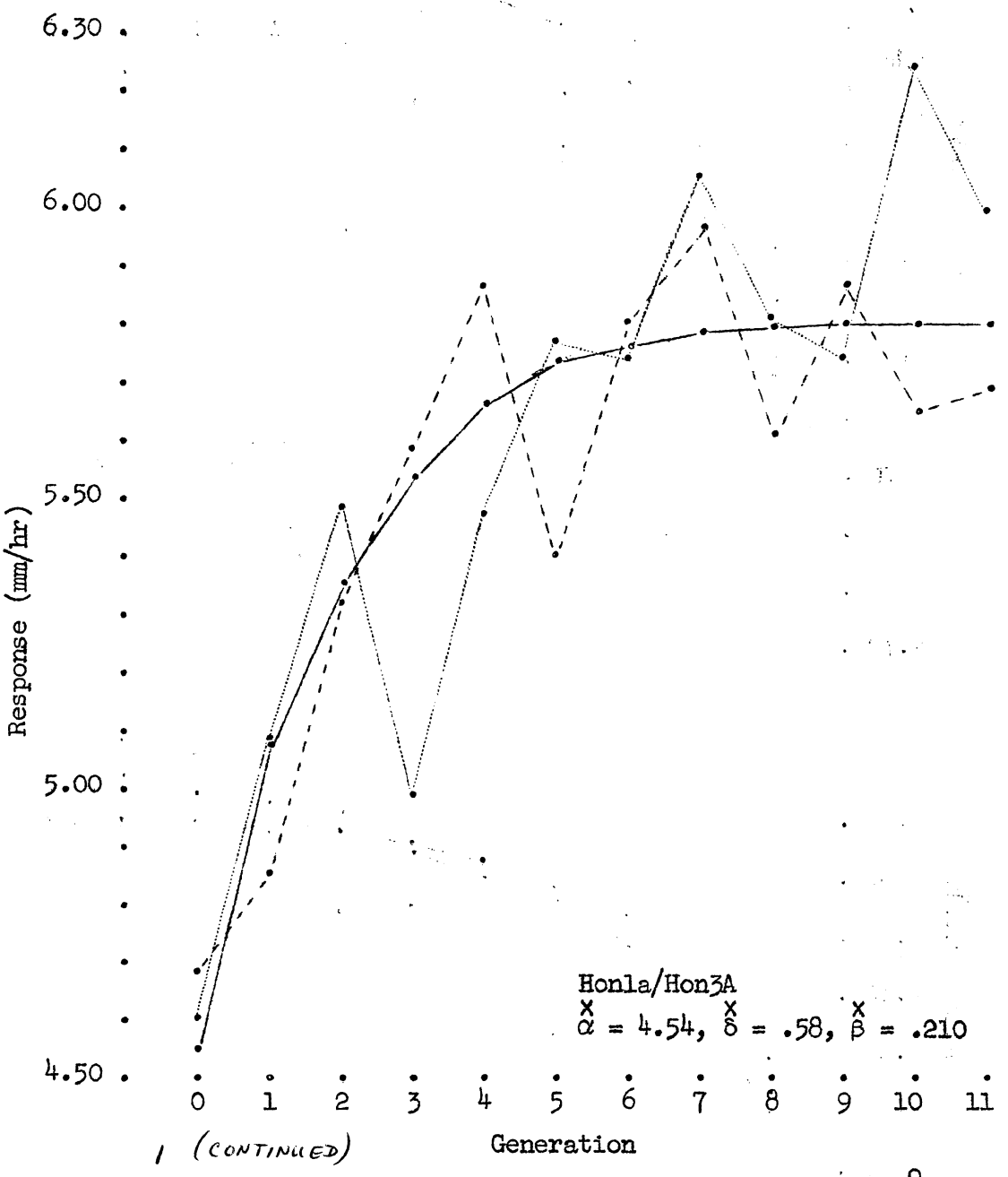


Figure 1 . Response to selection for fast growth rate at 18°C. in crosses Hon1a/Hon3A and Hon3A/Hon1a. Observed response for each replication is shown by broken lines, theoretical response shown by solid lines. Further explanations in the text.



1 (CONTINUED)

Figure 7. Response to selection for fast growth rate at 25°C. in crosses Honla/Hon3A and Hon3A/Honla. Observed response for each replication is shown by broken lines, theoretical response shown by solid lines. Further explanations in the text.



(CONTINUED)

Figure 3. Response to selection for fast growth rate at 35°C. in crosses Honla/Hon3A and Hon3A/Honla. Observed response for each replication shown by broken lines, theoretical response shown by solid lines. Further explanations in the text.

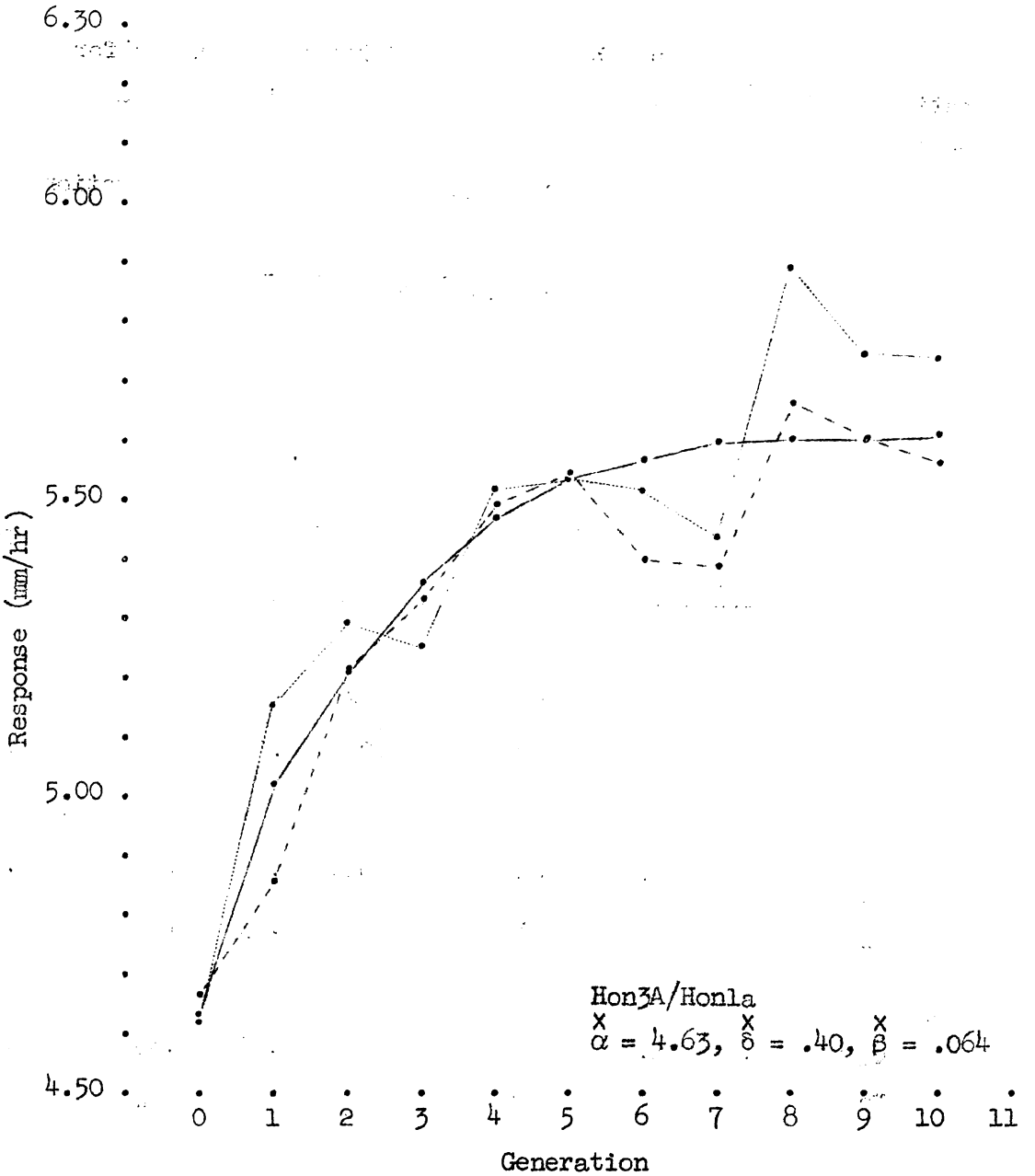


Figure 1. Continued

Table 1. Among and within individual selections mean squares for the Cross Honla/3A and its reciprocal, Hon3A/la for each of n cycles at each of three temperature environments, 18°C, 25°C, and 35°C.

Temp.	Cycle	Honla/3A				Hon3A/la			
		Replication I		Replication II		Replication I		Replication II	
		Among	Within	Among	Within	Among	Within	Among	Within
		m.s. (19 df)	m.s. (20 df)	m.s. (19 df)	m.s. (20 df)	m.s. (19 df)	m.s. (20 df)	m.s. (19 df)	m.s. (20 df)
18°C	0	.04053	.00597	.20343	.00115	.05463	.01555	.02939	.00301
	1	.00772	.00098	.01183	.00157[1]	.01791	.00368	.00588	.00050
	2	.01582	.00198	.01053	.00099	.00502	.00071	.00163	.00065
	3	.00456	.00063	.00507	.00109	.00558	.00281	.00346	.00268
	4	.00183	.00096	.01752	.00116	.00258	.00191	.00552	.00138
	5	.00121	.00106	.00386	.00221	.00311	.00078	.00220	.00170
	6	.00576	.00107	.00894	.00228	.00117	.00091	.00678	.00149
	7	.00181	.00041	.00395	.00125	.00217	.00132	.00938	.00177
	8	.00404	.00068	.00413	.00160	.01246	.00998	.01584	.00308
	9	.00142	.00080	.00292	.00322			.00302	.00134
	10	.00347	.00173	.00140	.00058			.00055	
	11	.00058		.00165					
25°C	0	.17712	.00423[2]	.30449	.00432	.07788	.03322[4]	.15085	.00831
	1	.03391	.00579	.01690	.00253[1]	.03375	.00494[2]	.01000	.00218
	2	.02335	.00247	.00611	.00217	.00687	.00530	.01104	.00315
	3	.02969	.00971	.00708	.00239	.01648	.00177	.02604	.00188
	4	.00523	.00221	.01651	.00324	.00390	.00662[1]	.03483	.00169
	5	.01024	.00241	.01039	.00242	.03099	.01532	.07372	.01545
	6	.01096	.00847	.05507	.00261	.01054	.00443	.00546	.00162
	7	.00605	.00390	.00736	.00194	.02433	.00678	.00767	.00457
	8	.00526	.00239	.00878	.00418	.01121	.00305	.01091	.00615
	9	.01202	.00345	.02131	.00598	.01495	.00452	.01150	.00714
	10	.01214	.00743	.00630	.00343	.00644	.00418	.00428	
	11	.00956	.00460[1]	.00656					
35°C	0	.64722	.00408	.59437	.00655	.39038[2]	.01019[3]	.37561	.01092[1]
	1	.07496	.00499	.04436	.00444	.06213	.00247	.03128	.00991
	2	.03557	.03957	.07917	.00546	.00575	.00382	.01903	.00349[1]
	3	.01272	.00214	.02188	.00646	.01274	.00292	.01565	.00542
	4	.01277	.00185	.06171	.00621[1]	.00429	.00204	.00449	.00139
	5	.00564	.00179	.00482	.00460	.01292	.00572	.03109	.00463
	6	.02277	.01314	.01357	.00693	.00970	.00331	.00757	.00461
	7	.00524	.00598[1]	.02462	.00975	.05667	.00447	.01121	.00266
	8	.00428	.00209	.12259	.01780	.00651	.00339	.01167[2]	.00388[1]
	9	.00867	.00349	.00969	.00847	.00611	.00291	.00919	.00246
	10	.01311	.00368	.00796	.00466	.00604	.00190[2]	.00215[2]	
	11	.00566		.00682					

[1] 19 df; [2] 18 df; [3] 16 df; [4] 13 df.

Table 2. Computations for $\hat{\alpha}$, $\hat{\delta}$, $\hat{W}_i$, $\hat{y}_i$ and mean square deviations from regression for both replications of the 18°C. data from cross Honla/Hon3A. $\hat{\beta} = .102$.

Cycle	$\hat{W}_i$	$\bar{x}_i$	$\bar{x}_i - (\Sigma\bar{x}_i/12)$	$\hat{y}_i$	$(\bar{x}_{i_1} - \hat{y}_i)^2$	$(\bar{x}_{i_2} - \hat{y}_i)^2$
0	.00000	2.35	-.2275	2.34	.0009	.0016
1	.95264	2.50	-.1225	2.47	.0004	.0064
2	1.59710	2.57	-.0525	2.56	.0009	.0036
3	2.01847	2.51	-.1125	2.62	.0081	.0144
4	2.20737	2.68	.0575	2.65	.0049	.0001
5	2.36147	2.59	-.0325	2.67	.0121	.0016
6	2.44708	2.60	-.0225	2.68	.0324	.0009
7	2.49264	2.74	.1175	2.69	.0001	.0100
8	2.51622	2.70	.0775	2.69	.0004	.0009
9	2.52822	2.71	.0875	2.69	.0001	.0036
10	2.53428	2.75	.1275	2.69	.0036	.0036
11	2.53732	2.77	.1475	2.69	.0196	0001
	24.19282	31.47	.0000		.0835	.0468

$$\sum_{i=0}^{11} \hat{W}_i^2 - \left(\sum \hat{W}_i \right)^2 / 12 = 6.99241821 \qquad \sum_{i=0}^{11} \left(\bar{x}_{i_1} - \hat{y}_i \right)^2 / 11 = .0076$$

$$\sum_{i=0}^{11} \hat{W}_i^2 \left[\bar{x}_i - \left(\sum_i \bar{x}_i / 12 \right) \right] = .97397835 \qquad \sum_{i=0}^{11} \left(\bar{x}_{i_2} - \hat{y}_i \right)^2 / 11 = .0043$$

$$\hat{\delta} = \frac{.97397835}{6.99241821} = .1393 \qquad \hat{\alpha} = \frac{\Sigma \bar{x}_i}{12} - \hat{\delta} \left(\frac{\Sigma W_i}{12} \right) = 2.34$$

$$\hat{y}_i = \hat{\alpha} + \hat{\delta}\, \hat{W}_i$$

Table 3. Predicted growth rates (mm/hr) for each cycle of selection in crosses Honla/Hon3A and Hon3A/Honla at 18°, 25° and 35°C. Predictions are based on the equations $\hat{y}_i = \hat{\alpha} + \hat{\delta}\,\hat{W}_i$ for the data from both replicates.

	Honla/Hon3A			Hon3A/Honla			Crosses combined		
Cycle	18°	25°	35°	18°	25°	35°	18°	25°	35°
0	2.34	4.09	4.55	2.35	3.91	4.65	2.35	2.98	4.58
1	2.47	4.27	5.04	2.49	4.10	4.99	2.48	4.19	5.02
2	2.56	4.40	5.35	2.57	4.22	5.22	2.57	4.33	5.31
3	2.62	4.47	5.55	2.62	4.30	5.38	2.63	4.41	5.49
4	2.65	4.52	5.56	2.65	4.33	5.48	2.66	4.44	5.59
5	2.67	4.54	5.73	2.66	4.36	5.54	2.67	4.47	5.65
6	2.68	4.56	5.76	2.67	4.37	5.57	2.68	4.48	5.68
7	2.69	4.57	5.78	2.67	4.38	5.58	2.69	4.49	5.69
8	2.69	4.57	5.79	2.67	4.39	5.59	2.69	4.49	5.70
9	2.69	4.57	5.79	2.67	4.39	5.60	2.69	4.49	5.70
10	2.69	4.57	5.80	2.67	4.39	5.60	2.69	4.49	5.71
11	2.69	4.57	5.80	-	-	-	2.69	4.49	5.71

Table 4. Mean square deviations from regression for both replications of the crosses Honla/Hon3A and Hon3A/Honla at 18°, 25° and 35°C. Regression analyses were performed on the mean of reciprocal crosses using various multiples of $\hat{\beta}$.

Multiple of $\hat{\beta}$	Honla/Hon3A		Hon3A/Honla	
	Rep. I	Rep. II	Rep. I	Rep. II
18°				
0	.00683$^-$	.00347$^-$	.00445$^+$	.00756$^-$
1/16	.00729	.00360	.00413	.00773
1/8	.00744	.00370	.00399	.00784
1/4	.00760	.00385	.00385	.00800
1/2	.00778	.00401	.00373	.00821
1	.00796	.00419	.00362	.00845
3/2	.00807	.00428	.00357	.00861
2	.00815	.00434	.00354	.00872
3	.00826	.00442	.00350	.00887
4	.00833	.00447	.00348	.00897
6	.00843	.00453	.00346	.00910
8	.00849	.00456	.00345	.00918
12	.00856	.00460	.00344	.00927
16	.00860$^+$	.00462$^+$	.00343$^-$	.00933$^+$
25°				
0	.02097$^-$	.02005$^+$	.01082$^-$	.02587$^-$
1/16	.02196	.01930	.01154	.02723
1/8	.02222	.01891	.01189	.02786
1/4	.02246	.01843	01235	.02859
1/2	.02269	.01790	.01291	.02934
1	02291	.01738	.01355	.03005
3/2	02303	.01710	01393	.03041
2	.02311	.01692	01420	.03064
3	.02322	.01670	.01456	.03093
4	.02329	.01657	.01478	.03110
6	.02338	.01642	.01506	.03130
8	02343	.01633	.01523	.03141
12	.02349	.01624	.01542	.03153
16	.02352$^+$	.01619$^-$	.01553$^+$	.03160$^+$
35°				
0	.06084$^-$	.03262	.01840$^-$	.02683$^-$
1/16	.06441	.03032	.01982	.02736
1/8	.06625	.02965	.02037	.02745
1/4	.06868	.02906	.02098	.02757
1/2	.07163	.02874$^-$	.02163	.02782
1	.07487	.02888	.02230	.02831
3/2	.07675	.02923	.02268	.02873
2	.07804	.02958	.02294	.02908
3	.07973	.03020	.02327	.02964
4	.08082	.03069	.02349	.03004
6	.08216	.03140	.02377	.03060
8	.08297	.03188	.02394	.03097
12	.08392	.03248	.02414	.03143
16	.08446$^+$	.03284$^+$	.02426$^+$	.03170$^+$

+, - equal maximum and minimum values, respectively